The future of string theory

John H. Schwarz[a)]
California Institute of Technology, Pasadena, California 91125

(Received 2 January 2001; accepted for publication 13 February 2001)

Prophesy is just for fun. The more useful purpose of the exercise is to identify important issues and to stimulate thought about where they stand and how they might be resolved. The subject areas that are fair game include all of particle physics and cosmology.
[DOI: 10.1063/1.1377276]

I. THIRTY YEARS OF PROGRESS

Since my topic is the future of string theory, I will discuss issues and problems that are currently unresolved and where there is still confusion, doubt, and uncertainty. However, if I were to do only that, it would give a distorted view of a subject that has undergone tremendous progress over the past thirty years. Therefore, to give a more balanced picture, I have decided to begin by presenting a brief chronology of some of the major developments that have taken place. I think we are entitled to look back on this with considerable pride and satisfaction.

- 1968–70: String theory developed to describe the strong interactions (hadron physics)
- 1971–73: Supersymmetry introduced in string theory and field theory
- 1974: String theory reinterpreted as a unified theory of gravity and other forces
- 1976–78: Supergravity; Montonen–Olive duality
- 1977–83: Superstring theory; path-integral formulation
- 1984: Anomaly cancellation; heterotic strings; Calabi–Yau compactification
- 1985–89: Conformal field theory; superstring perturbation theory; T duality; mirror symmetry; string field theory; matrix models
- 1990–94: S duality; p-branes; Seiberg–Witten theory
- 1995: Topology change; M theory; dualities relating all superstring theories and M theory; D-branes
- 1996: Black-hole entropy; F theory; matrix theory
- 1997–99: Brane configurations; AdS/CFT duality; noncommutative geometry; brane worlds

II. THE ISSUES

Let me now discuss where I think future progress will occur. There are two major fields that string theory ought to illuminate some day: particle physics and cosmology. (A third one is mathematics, but I will not discuss that.) Even though we think of particle physics as addressing the microscopic extreme and cosmology the macroscopic extreme, it is quite natural to consider them together when dealing with a theory that incorporates gravity, as string theory does. Relativists and particle theorists have both identified the important problem of reconciling quantum theory with general relativity. The prospect of achieving this attracts both of them to string theory. It is clear to many of us that string theory really does achieve this reconciliation, but it is also clear that there are important aspects of how this works that are not yet fully understood.

[a)]Electronic mail: jhs@theory.caltech.edu

One such issue concerns the status of quantum mechanics and information loss. In 1976, Hawking argued that the existence of black holes implies that a pure quantum state can evolve into a mixed state, in other words there is a loss of quantum coherence. Were this to happen, it would mean a breakdown of quantum mechanics. I think it is quite clear that string theory respects quantum mechanics, and it might even explain it. It gives unitary evolution without loss of coherence, and it can also describe black holes. Thus, it should be possible to study black holes in string theory and to explain precisely how Hawking's argument breaks down. Despite a great deal of effort, I don't think that this has yet been done in a clear and convincing way. However, I certainly believe that it should be possible.

Let me now turn to the most vexing and far-reaching of the unresolved issues: the cosmological constant.

A. The cosmological constant

The low energy effective theory of gravity contains the standard Einstein–Hilbert term:

$$S_{\mathrm{E}}=\frac{1}{16\pi G}\int \sqrt{-g}R\, d^4x.$$

Theoretically, it is natural to also include a vacuum energy, or cosmological, term,

$$S_\Lambda=\int \Lambda\sqrt{-g}\, d^4x.$$

Here, the parameter Λ—called the cosmological constant—can be interpreted as the energy density of the vacuum. Such a term is generically induced by radiative corrections even if it is zero in the classical theory. In particular, it receives contributions from the zero-point energies of all fields in the theory. You need a well-defined quantum theory of gravity, before the vacuum energy becomes something computable. String theory is the only such theory on the market. So it is only with the advent of string theory that proposals for understanding this parameter from a fundamental theoretical viewpoint can be analyzed.

Observationally, the bounds on Λ are exceedingly severe. As a first approximation, one can simply require that the energy in the form of vacuum energy is less than the amount that would give closure of the universe. Using this bound, one finds that Λ, expressed in Planck units, is less than 10^{-122}. This is the best measured approximation to zero of any physical parameter in nature. We do not yet have a convincing way to derive why Λ should be this small in the context of a realistic model. Even so, it seems much more likely that someday we will be able to derive the answer zero than an answer of order 10^{-122}.

In recent years, astrophysicists and cosmologists have settled on a fairly precise inventory of how mass and energy are distributed in the universe. Important constraints come from studies of the cosmic microwave background, large scale structure, and Type Ia supernovas, together with the rest of standard cosmology. Based on this, they have concluded the following: The total mass and energy in the universe give, to within about 10%, the critical closure density. This is the value required by inflation, which gives a flat open universe. It has been clear for some time that an appreciable fraction of the mass of the universe is in the form of *dark matter*, but now it appears that there is a second mysterious component, which could be called *dark energy*. The inventory of mass and energy is roughly as follows: 5% is baryonic matter, 25% is dark matter (mostly cold dark matter, but a small portion could be massive neutrinos), and 70% is dark energy. Cold dark matter is concentrated in the halos of galaxies, whereas dark energy is distributed uniformly throughout the universe.

Each component has an equation of state of the form $p=w\rho$, where p is pressure and ρ is density. Nonrelativistic matter has $w=0$ and radiation has $w=1/3$. To fit the data, the dark energy should have an equation of state with $-1.0\leqslant w<-0.5$. Future observations should determine this parameter with greater accuracy. The most popular guess, which fits the data well, is that the dark energy is vacuum energy (i.e., a cosmological constant). This gives $w=-1$.

Components of the universe with different values of w evolve differently:

$$\rho \sim R^{-3(1+w)},$$

where R is the scale factor of the universe. Thus the universe was primarily comprised of matter (and radiation) in the distant past and will be more and more comprised of dark energy in the future. This means that the present epoch of the universe is rather special: it is the epoch in which the contributions of matter and dark energy are comparable. This is a surprising coincidence, but it seems to be a fact.

As we have already indicated, if the dark energy is vacuum energy, this corresponds to a value of the cosmological constant that is so small that it seems implausible that we will be able to derive it. Therefore it behooves us to ask whether there are any other candidates for the dark energy.

One alternative to a cosmological constant, which has been discussed quite a bit lately, goes by the name of *quintessence*—a word that originally referred to the human soul. In current usage it is the energy carried by a scalar field, which is not at the minimum of the potential, but is still rolling towards its minimum value. If it rolls sufficiently slowly, the phenomenology becomes very similar to that of a cosmological constant. (This kind of slow-roll mechanism is reminiscent of ''new inflation,'' though at a vastly lower energy scale than in the case of the big bang.) In such a scenario the fundamental value of Λ could be zero, if that is the minimum of the potential that is being approached asymptotically. In this way one avoids the need to explain an exceedingly small cosmological constant, but other equally challenging puzzles arise. For one thing, getting the field to roll slowly enough involves a fine-tuning of parameters that is just about as formidable as that for a tiny cosmological constant.

Such a slowly rolling scalar would correspond to a spin zero particle that is essentially massless, since the curvature of the potential would correspond to a Compton wavelength comparable to the size of the universe. Therefore this scalar could mediate long-range scalar forces. There are many ways in which massless scalar fields can arise in string theory, such as moduli associated with compactification and a dilaton. All of them typically have couplings to ordinary matter that is roughly of gravitational strength. The fact that gravity is observed to be a purely tensor force, to better that one percent accuracy, severely restricts the possibilities for massless scalars. So this seems to me to be a problem for accommodating the quintessence proposal in the context of string theory.

Another issue to be considered is that the values of scalar fields in string theory control observable parameters such as Newton's constant and the fine structure constant. A rolling scalar would therefore be expected to imply that these are changing with time. Observational bounds on such time variation are quite severe, so this is also a problem for the quintessence proposal.

If neither a cosmological constant nor quintessence is the right answer, what other possibilities are there? There has been some discussion of schemes in which the dark energy would be carried by topological defects. For example, domain walls that are solid (i.e., they resist shear) and have certain other special properties could give $w=-2/3$. As it stands, this does not look very convincing. However, neither do any of the alternatives.

So, to conclude this part of the discussion, even if we knew how to prove that the cosmological constant vanishes in string theory, there would still be a serious problem accounting for the cosmological observations. It will be interesting to see whether future high-precision determinations of the cosmological parameters confirm the present picture.

B. What is the role of supersymmetry?

Supersymmetry appears to be an essential feature in string theory and M-theory that is required to ensure mathematical consistency. Therefore it seems pretty clear that supersymmetry should be physically relevant at the fundamental scale, which is either the string scale or the eleven-dimensional Planck scale. But a crucial question, whose answer is much less certain, is whether supersymmetry is also relevant to the description of physics at the electroweak scale. There are several unrelated arguments that suggest that a supersymmetry is broken near the electroweak scale. Even though none of them is conclusive, it is very impressive that they each lead to roughly the same scale for the typical mass of superpartners.

(1) Supersymmetry provides a solution to the gauge hierarchy problem. The ratio of the electroweak scale to the unification scale or the string scale is around 10^{-14}. In the context of the standard model this is puzzling, since quadratic divergences in the Higgs mass would renormalize the Higgs mass (and hence the electroweak scale) up to the cutoff. Supersymmetry builds in cancellations that softens these divergences to being only logarithmic, thereby protecting the hierarchy from being destroyed by radiative corrections. (This does not explain where the hierarchy comes from in the first place.) This reasoning requires that the energy scale that characterizes supersymmetry breaking should be comparable to the electroweak scale—i.e., around 100 GeV to 1 TeV. This argument is not a proof of supersymmetry, because there could be other solutions to the hierarchy problem. One is known, and there could be others.

(2) The unification of the three gauge couplings at a high energy scale works much better with supersymmetry than without it. In fact, studying the fits with a variable supersymmetry breaking scale one finds that unification of the couplings is achieved for a supersymmetry mass gap less than a few TeV. This is very impressive, but one could certainly imagine that other new physics at intermediate scales could also lead to successful unification.

(3) A neutralino LSP with a mass of about 50–500 GeV is an excellent candidate for cold dark matter. It is not possible to be much more precise than this, because we do not know yet the mixture of gauginos and Higgsinos that makes up the LSP, and that affects the relationship between the LSP mass and the cosmological mass fraction that it provides. If the LSP is the dominant component of cold dark matter, the universe could have five times as much mass in neutralinos as in baryons.

(4) In the context of supersymmetric grand unified models the renormalization group running of Higgs masses can give rise to electroweak symmetry breaking at roughly the right scale. In this way supersymmetry helps in establishing the hierarchy as well as in protecting it from radiative corrections.

Supersymmetry may be relevant to solving the cosmological constant problem. There are many known string theory solutions that give a flat Minkowski spacetime with unbroken supersymmetry. They are unrealistic, of course, since supersymmetry has to be broken. However, for these solutions, supersymmetry ensures that radiative corrections do not generate a cosmological constant. So, at least in the context of unrealistic solutions, there is a symmetry explanation of the vanishing of the cosmological constant. Unfortunately, it seems that this cancellation mechanism only works for solutions with unbroken supersymmetry. When supersymmetry is broken, one expects to get a cosmological constant with a size controlled by the supersymmetry mass gap. Assuming that scale is around 1 TeV, the resulting cosmological constant is still some 56 orders of magnitude too big. That's an improvement on 124 orders of magnitude, but it leaves a lot of room for further progress.

Therefore, I like to pose the problem of the cosmological constant as the following question: *Is there a supersymmetry breaking mechanism that does not generate a cosmological constant?* Such a mechanism is not known, but it seems to me that this is what we need. If one were discovered, that would be very exciting. Generic breaking of supersymmetry in the MSSM introduces over 100 new parameters, which is one of the reasons that it is so difficult to use it to make quantitative predictions. I would expect the ''right'' supersymmetry breaking mechanism to be highly constrained and therefore much more predictive.

Lest I leave you with the wrong impression, I should point out that this problem has received a lot of attention over the years. Ingenious proposals have been put forward by Moore, Witten, Kachru and Silverstein, and others. However, I suspect that the correct solution still remains to be discovered. It might be that insights gained from AdS/CFT duality will provide a better framework for thinking about the cosmological constant. In any case, the importance of this problem cannot be overstated.

C. What is the theory?

What is the theory? What is the principle on which it is based? What is the best way to formulate it? It is rather striking that after 30 years of enormous progress and effort by many hundreds of the most talented physicists, string theory is not yet fully formulated.

Matrix theory and AdS/CFT duality can be viewed as providing exact nonperturbative definitions of string theory or M-theory for certain classes of solutions. This is a remarkable achievement, but we want more than a new recipe for each solution. We want a single formulation that applies to all possible solutions. When such a formulation is found, it is entirely possible that the name *string theory* will no longer be considered appropriate. I suspect that *M-theory* will not fill the bill either. But this is a secondary issue. Understanding the theory is much more important than naming it.

I do not know what form the theory will take when it is completed. It might, for example, be based on some abstract algebraic structure. The concepts of space and time are likely to emerge as properties of particular solutions rather than as smooth background geometries on which the theory is formulated in the first place. It is hard to begin to formulate a theory without reference to spacetime, since it is so radically different from anything we have dealt with before. The standard recipe would have us introduce quantum fields for particles or strings or whatever, and then formulate an action that describes their propagation in a given spacetime manifold. What we seem to need is a theory for which the particles or the strings, as well as the spacetime manifold, are properties of particular solutions rather than features of the underlying theory itself.

I expect that the optimal formulation of the theory will eventually be found, but I would not wish to attach a time frame to this prediction. One lesson I have learned during my career is that it is very hard to anticipate future developments. String theory has undergone several revolutions that have profoundly changed the way we think about the subject. Each one has caused us to ask new questions that we would not have even posed before. It is not at all clear to me how many more such revolutions are still required before we are in a position to formulate the theory properly. I would guess that the process is finite, and that we will eventually get there, but it is hard to assess how close we are at the present time. Once found, this theory will surely be a thing of great beauty, based on profound physical and mathematical principles. As I mentioned earlier, it will probably be sufficiently different from the present formulations to justify giving the subject a new name. However, I do not think there is yet a good reason to drop the name ‘‘string theory.’’

D. What is the right solution?

Formulating the theory is not the whole problem. What is likely to prove to be even harder is the determination of a realistic solution. While I am convinced that the theory should be completely unique, with no adjustable parameters, I suspect that the story for solutions is very different. Every indication is that the theory admits a large number of different solutions, most of which are completely unrealistic. Indeed, the ones that are known are supersymmetric, and therefore certainly unrealistic.

There are a number of obvious questions that need to be answered: Can one completely classify all solutions of the theory? Is there a realistic solution with a Lorentz invariant vacuum, that gives an accurate description of particle physics? Is there a cosmological solution that describes the evolving universe that we observe? Can these two questions be addressed separately or do we need to understand cosmology to give a proper description of particle physics?

For the reasons I discussed earlier, it would probably be better if there were no massless or rolling scalars. An added bonus of this is that it would mean that the quantum effective potential that describes the dynamics of all light scalar fields has an isolated minimum. If such a minimum could be determined theoretically, the corresponding solution would have no continuously adjustable parameters, and all physical quantities would become computable in principle. This would be a very satisfying outcome.

E. Extra dimensions of space

Besides fermionic dimensions, string theory also requires extra spatial dimensions of a somewhat more conventional type. These have generally been assumed to form a compact space of six or seven dimensions with a size that is roughly comparable to the string scale, which might be some 10^{-32} cm. The details of the geometry of the compact space profoundly influence the physics that is observed in four large dimensions. Thus we can hope to infer the geometry indirectly, even though such small extra dimensions would not be directly accessible by foreseeable experiments.

Some physicists have recently suggested that the extra dimensions could be much larger than previously envisioned, maybe even so large as to be experimentally observable. There is now a rather large community actively exploring possible scenarios of this type. To be honest, I find the idea of large extra dimensions to be rather implausible, since it undermines some of the successful predictions of grand unification, such as unification of the couplings and suppression of proton decay. Nevertheless, I do feel that it is worthwhile to explore these ideas, and to improve the experimental bounds. Who knows what might turn up?

F. The role of experiment

I do not expect the right solution to be found by pure thought alone. Experimental guidance, and the more traditional back and forth between theory and experiment, will be important in this quest. One central issue that experimentalists will settle in the next ten or twenty years is what the new physics is that is responsible for electroweak symmetry breaking and mass generation.

The experiments will surely discover one or more Higgs particles. The bigger question, in my opinion, is whether they will find evidence for supersymmetry that is broken at the electroweak scale. If this is the case, then supersymmetry particles should be discovered experimentally at Fermilab or CERN in the early part of the century. This discovery would have profound theoretical and experimental consequences. It would provide evidence for a nontrivial extension of the known symmetries of space and time, an extension which could be described as the discovery of fermionic dimensions. Also, it would set the agenda for experimental particle physics for several decades. When I want to be sure to be quoted, I tell reporters that the discovery of supersymmetry would be more profound than life on Mars.

In addition to experiments at particle accelerators, there are also important nonaccelerator experiments being carried out. For example, dark matter searches are underway. Over the next few years they are expected to achieve the sensitivity required to cover most of the parameter space that is favored for axions or neutralinos. My guess is that a neutralino LSP will eventually turn up in these searches.

Such discoveries would make it clear that the abstract mathematical musings of the past thirty years can be connected to experimental science. Experimental facts about supersymmetry at the electroweak scale would provide crucial guidance in the quest to understand how to connect the underlying theory to the real world. Maybe it would even guide us to the discovery of a supersymmetry breaking mechanism that does not generate a cosmological constant.

G. The role of communication

The final decade of the old millennium saw the birth—from within particle physics—of a powerful new technology called the World Wide Web, which has been quietly revolutionizing education and bringing young people closer to physics than they could ever get when I was a boy in school. For example, on the Superstringtheory.com web site over Christmas break, the following message appeared posted by a teenager from Canada:

'This was a great site that agreed with many of my own theories of ''strings.'' For the first time in my life it seems that someone seems to know what the heck I ''know.'' (I am in grade 9 so when I bring up the topic of quantum mechanics I usually get a fuzzy stare.) Keep up the excellent work.'

So not only is the next generation of string theorists out there, we have a way to reach them, and they have a way to reach us, that has never existed before in history.

Therefore I would like to predict that the future could look very bright in that direction, if we take full advantage of this new technology.

III. CONCLUSION

In conclusion, string theory has developed into one of the most active areas of theoretical physics in recent years. The last third of the 20th century witnessed the construction of an amazing mathematical edifice, which we are struggling to understand. I expect to see equally striking progress achieved in the first third of the 21st century.

The theory is still not fully understood, but I am optimistic that a deeper formulation, which makes clear that it is a unique theory, will be found. I also think that there is a good chance that we will understand how to find solutions that are able to account for the observations of particle physics and cosmology.

The construction of the standard model arose as a collaborative effort in which theorists and experimentalists both made major contributions. In the case of string theory, the natural energy scale is much higher, and so it is more difficult to make contact with experiment. To date, the interactions with mathematicians have played a bigger role than the ones with experimentalists. In the future, I expect that both will be very important. The subject involves a lot of new and bizarre concepts, much as quantum mechanics did in the first half of the 20th century. Before we are through, it is likely that more of them will be identified. To get it right, we will need the help of our mathematical and experimental friends. It is only fitting that a theory that unifies particles and forces should also unify disciplines.

ACKNOWLEDGMENTS

This work work was supported in part by the U.S. Department of Energy under Grant No. DE-FG03-92-ER40701.